



A soluble, folded protein without charged amino acid residues

Højgaard, Casper; Kofoed, Christian; Espersen, Roall; Johansson, Kristoffer Enøe; Villa, Mara; Willemoës, Martin; Lindorff-Larsen, Kresten; Teilum, Kaare; Winther, Jakob R.

Published in:
Biochemistry

DOI:
[10.1021/acs.biochem.6b00269](https://doi.org/10.1021/acs.biochem.6b00269)

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Document license:
[Other](#)

Citation for published version (APA):
Højgaard, C., Kofoed, C., Espersen, R., Johansson, K. E., Villa, M., Willemoës, M., Lindorff-Larsen, K., Teilum, K., & Winther, J. R. (2016). A soluble, folded protein without charged amino acid residues. *Biochemistry*, 55(28), 3949–3956. <https://doi.org/10.1021/acs.biochem.6b00269>

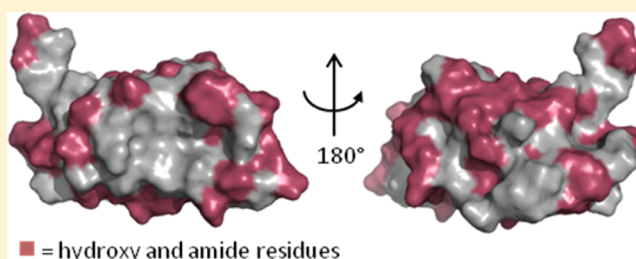
A Soluble, Folded Protein without Charged Amino Acid Residues

Casper Højgaard, Christian Kofoed,[†] Roall Espersen,[‡] Kristoffer Enøe Johansson,[§] Mara Villa,^{||} Martin Willemoës, Kresten Lindorff-Larsen, Kaare Teilum, and Jakob R. Winther*

Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

Supporting Information

ABSTRACT: Charges are considered an integral part of protein structure and function, enhancing solubility and providing specificity in molecular interactions. We wished to investigate whether charged amino acids are indeed required for protein biogenesis and whether a protein completely free of titratable side chains can maintain solubility, stability, and function. As a model, we used a cellulose-binding domain from *Cellulomonas fimi*, which, among proteins of more than 100 amino acids, presently is the least charged in the Protein Data Bank, with a total of only four titratable residues. We find that the protein shows a surprising resilience toward extremes of pH, demonstrating stability and function (cellulose binding) in the pH range from 2 to 11. To ask whether the four charged residues present were required for these properties of this protein, we altered them to nontitratable ones. Remarkably, this chargeless protein is produced reasonably well in *Escherichia coli*, retains its stable three-dimensional structure, and is still capable of strong cellulose binding. To further deprive this protein of charges, we removed the N-terminal charge by acetylation and studied the protein at pH 2, where the C-terminus is effectively protonated. Under these conditions, the protein retains its function and proved to be both soluble and have a reversible folding–unfolding transition. To the best of our knowledge, this is the first time a soluble, functional protein with no titratable side chains has been produced.



Charge is one of the fundamental properties thought to be required for proper folding and function of proteins, and protein charges are considered to be crucial for protein solubility. Seven different amino acid residues (Asp, Glu, Arg, Lys, His, Cys, and Tyr) have titratable side chains in a pH interval from 2 to 12. Of these, the first five are typically categorized as the charged residues^a, because they are usually found in an ionized state near neutral pH. In an average, soluble protein, Asp, Glu, Arg, Lys, and His together comprise ~25% by number.^b In accordance with their high prevalence, it is generally accepted that surface charges provide a means for specificity and topological orientation of surface interactions in proteins.^{1,2} At the same time, charge–charge interactions are generally regarded as being important to protein solubility by inducing long-range repulsion between alike-charged species. This has been clearly demonstrated by introduction of charged side chains on a massive scale, termed “supercharging”, and has, albeit at a cost of thermodynamic stability, in several instances been shown to provide an increased resilience toward aggregation, particularly at elevated temperatures.^{3–5} In line with this observation, thermo-stable proteins, in general, contain more surface charges than their mesophilic counterparts.^{6–8} Neutralization of surface charges is suggested to be required for efficient amyloid fibril formation,⁹ and chemical and mutational neutralization of charges has indeed been shown to promote the rate of fibrillation.^{10–12}

A better understanding of how surface charges affect protein solubility and aggregation rates has many important medical and industrial applications.¹³ It has been shown that charged amino acids contribute favorably to solubility in a pH- and context-dependent manner.^{14,15} However, the solubility of native proteins appears to correlate only with the fraction of the accessible surface area that is negatively charged, not with the fraction that is positively charged.¹⁶ This has been suggested to be due to weak hydration of arginine and lysine residues.^{14,17}

The effects of salt bridges on protein stability have been studied using double-mutant cycles,¹⁸ and although the significance of naturally occurring salt bridges remains controversial,^{19,20} stabilization through protein engineering has been reported.^{21,22} In any case, the stabilizing effects of individual salt bridges are highly context-dependent, as they depend on the accessible surface area of the salt bridge,²³ long-range charge–charge interactions with the rest of the protein,²⁴ and the specific geometry of residues involved.²⁵ In addition to local interactions such as salt bridges, Coulombic interactions between charged residues distinguish themselves by being able to act over relatively long distances. Thus, a change in the charge state of one residue, when titrated, may influence many other titratable residues, resulting in a complex overall titration

Received: March 24, 2016

Revised: June 3, 2016

Published: June 16, 2016

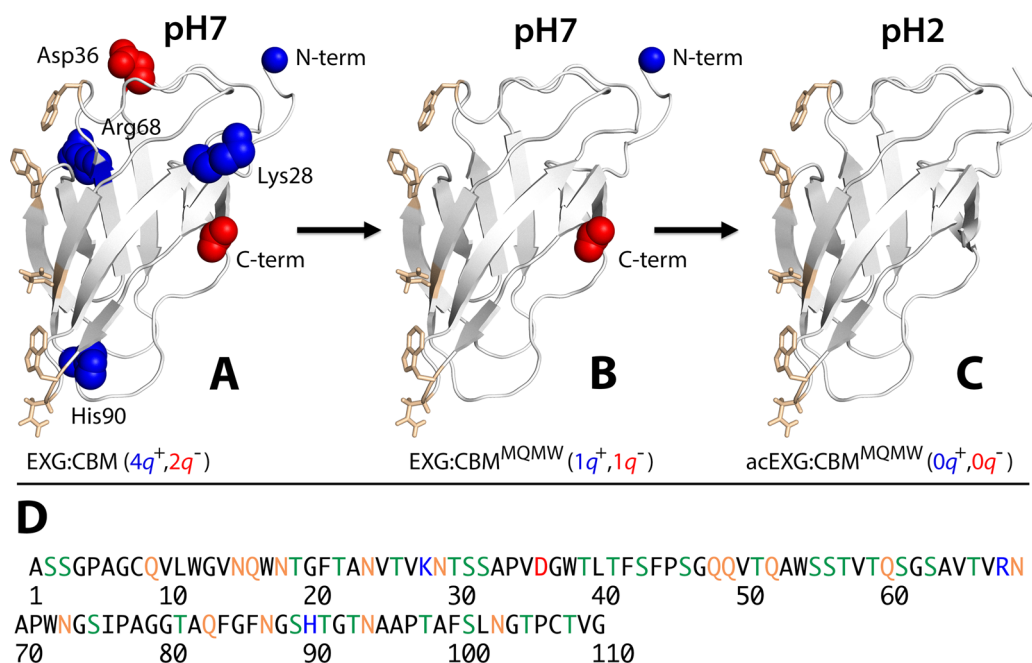


Figure 1. Charge depletion of EXG:CBM. (A) Wild-type EXG:CBM has a potential of four positive (blue) and two negative (red) charges at neutral pH (assuming the “standard” pK_a values for all residues). None of the residues are located within the cellulose-binding site, which includes residues Trp17, Trp54, Trp72, Asn15, and Asn87 (represented as sand-colored sticks). (B) The EXG:CBM^{MQMW} mutant has the potential for two charges, because of only its two termini. (C) Acetylation of the N-terminus renders the protein acEXG:CBM^{MQMW}, which by protonation of the C-terminus at pH 2.0 becomes completely uncharged. (D) Sequence of EXG:CBM. The mutated residues are colored blue (positive residues) and red (negative residue). Hydroxy amino acids are colored green, and amide amino acids are colored orange.

behavior that is difficult to follow experimentally^{26,27} and predict computationally.^{28,29}

To evaluate the relevance of bulk surface charge, attempts have been made to produce a protein without titratable residues. Studies have shown both ubiquitin and ribosomal protein S6 to be folded under conditions where they are essentially uncharged,^{30,31} demonstrating that surface charges are not essential for protein stability or *in vitro* folding. To make ubiquitin uncharged, a combination of mutagenesis, bulk chemical modification, and pH titration was used. The construction of charge-free S6 followed a different approach, in which all the positively charged residues were replaced with serines and the negative charges titrated. Although this was attempted, it was not possible to produce charge-free S6 by mutagenesis alone. Both uncharged proteins differed from their parent proteins, in that they suffered from low solubility. Further, while both proteins are essentially uncharged at low pH, they become highly charged at higher pH values.

It is generally thought that charged, hydrophobic, and polar residues are necessary to build globular folded proteins. In line with this, the attempts to construct “reduced alphabet” proteins have always included representative members of all three groups to cover the biophysical properties deemed necessary for protein function.^{32–35} The charge-free proteins mentioned above^{30,31} challenge this view by being both stable and folded. However, in both cases, the proteins still contained carboxylic amino acids and the charge depletion came at a cost in the form of low solubility. To make further progress in making a charge-free protein, we chose to start from a protein already carrying only a few charges and to apply a reduced amino acid alphabet of 14 amino acids in which the class that is able to carry a charge in the given protein context has been completely omitted. This approach made us ask the simple question: is it at

all possible to produce *in vivo* a protein without charged amino acids that is both soluble and functional? Indeed, we have produced the first known example, artificial or natural, of a fully functional and highly soluble protein without charged amino acids. The chargeless protein is expressed well in *Escherichia coli*, retains high solubility, and maintains its biological function in the complete pH range from 2 to 12.

EXPERIMENTAL PROCEDURES

Design of a Charge-Free Protein. Residues at positions 26, 36, and 68 are exposed to solvent, while the histidine at position 90 has only limited surface exposure. To identify the least destabilizing mutations, we calculated the change in free energy of unfolding for all noncharged substitutions using FoldX.³⁶ We modeled the four positions independently justified by the long interpositional distance in the three-dimensional (3D) structure (Figure 1). By combining the results from slightly different protocols for structural relaxation, we identified a number of potential substitutions at each site where the mutations would either stabilize the protein or result in only a small decrease in stability. In combination with a visual inspection of models of the mutant variants, we selected the quadruple mutant K28M/D36Q/R68M/H90M (EXG:CBM^{MQMW}) for our initial studies. As H90 potentially was the most challenging position, we also expressed the K28M/D36Q/R68M triple mutant and compared its stability to that of the protein in which all four charged residues were substituted. This comparison indeed showed that the H90M mutation was more deleterious than the three others as both expression levels and stability were compromised in the quadruple mutant. This prompted us to investigate additional variations at position 90, mutations to W, F, S, T, N, Q, and V,

resulting in the final chargeless variant presented here, EXG:CBM^{MQMW}.

Expression and Purification. Protein was prepared as described elsewhere,³⁷ although EXG:CBM^{MQMW} eluted from Avicel (Sigma-Aldrich) at a lower temperature (68–75 °C). The yield of EXG:CBM^{MQMW} after the final purification (~2 mg/L) was also substantially lower than that of EXG:CBM (20 mg/L) in part because of a decreased resistance to lyophilization. Acetylation of EXG:CBM^{MQMW} was achieved by adding a 100-fold molar excess of acetic anhydride followed by chromatography on a Superdex 75 column. Complete acetylation was confirmed by mass spectrometry.

Cellulose Binding Assay. Two micrograms of each EXG:CBM variant (4 µg of acEXG:CBM^{MQMW}) and 2 µg of ribonuclease A (GE Healthcare) were diluted into 50 mM sodium phosphate buffer at the appropriate pH, and 1 mg of Avicel was added. The samples were incubated for 1 h at 4 °C, followed by centrifugation (5 min at 16000g). Supernatants were transferred to new Eppendorf tubes, and the cellulose pellets were washed twice with buffer. Supernatants and control samples without Avicel were precipitated by adding TCA to a final concentration of 12%. Cellulose and TCA pellets were resuspended directly in sample buffer and used for Tricine–sodium dodecyl sulfate–polyacrylamide gel electrophoresis (Tricine–SDS–PAGE) as described previously.³⁸

Solubility Measurements. For measurements of precipitation propensity, lyophilized protein was suspended in 10 mM sodium phosphate buffer at the appropriate pH and monitored after centrifugation for 20 min at 20000g followed by measurement of the absorbance at 280 nm on a NanoDrop ND-1000 spectrophotometer. Samples were left at room temperature between measurements. Extinction coefficients from ExPASy ProtParam³⁹ were used to convert absorbance to protein concentration. The solubility in ammonium sulfate was measured as described previously.¹⁴ All measurements were taken in triplicate.

Stability Measurements. Guanidine-induced unfolding was monitored via fluorescence spectroscopy using a PerkinElmer LS55 luminescence spectrometer. The samples [~0.3 µM protein with an appropriate amount of GdnHCl in 50 mM glycine (pH 2.6) or 50 mM MOPS (pH 7.0)] were excited at 280 nm, and the intensity of the emission at 350 nm was measured. All measurements were taken at 25 °C. Data were fitted using the linear-extrapolation method.⁵⁶ The signal did not change significantly between 1 and 24 h after mixing the samples, so 1 h was chosen for equilibration.

Circular Dichroism (CD). Changes in the secondary and tertiary structure as a response to pH were followed using CD signals in the far- and near-UV area (190–260 and 250–340 nm, respectively) with a J-810 Jasco spectropolarimeter. The proteins (10–50 µM) were diluted in 10 mM potassium phosphate buffer at the appropriate pH. All measurements were performed at 25 °C. Each spectrum shown is an average of 10 (far-UV) or 5 (near-UV) spectra.

NMR Spectroscopy. Cultures were grown in MOPS medium with ¹⁵NH₄Cl (98%, Sigma-Aldrich) as the nitrogen source. The proteins were expressed and purified with Avicel as for the unlabeled protein. The protein was dissolved in 90 mM KCl, 10% D₂O, 1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), and 0.02% NaN₃ to a final concentration of ~1.1 mM (EXG:CBM) or ~0.3 mM (EXG:CBM^{MQMW}). Protein solutions were brought to the starting pH and titrated with either NaOH (EXG:CBM) or HCl (EXG:CBM^{MQMW}). The

pH of the sample was determined before and after each NMR measurement and was generally in good agreement (<0.1 pH). Measurements were taken in Shigemi tubes on a 750 MHz (EXG:CBM) or 800 MHz (EXG:CBM^{MQMW}) Varian Inova spectrometer at 25 °C. The Varian BioPack gNhsqc.c pulse sequence was used for HSQC experiments and the sofastNhmqc.c pulse sequence for HMQC experiments. pK_a values were obtained by calculating a weighted chemical shift change from the proton (ΔH¹) and nitrogen (ΔN¹⁵) chemical shifts of each backbone amide using⁴⁰

$$W\Delta\delta = \sqrt{(\Delta H^1)^2 + (0.154\Delta N^{15})^2}$$

and fitting to the Henderson–Hasselbalch equation of the form

$$W\Delta\delta = \frac{W\delta_A 10^{pH-pK_a} + W\delta_{AH}}{1 + 10^{pH-pK_a}}$$

Only strong, unambiguous assigned chemical shifts were used for the calculation (resulting in different cutoffs in the basic region of the pH scale for the different residues), and only signals with significant changes ($W\Delta\delta_{\max} \geq 0.1$) were considered. The pK_a values of the C-terminus and Asp36 were determined directly from their respective amid signals. Because of line broadening and loss of the His90 amid signal, the pK_a value of His90 was determined from a global fit of nine of the spatially closest residues (S46, G47, Q48, Q49, G88, S89, Thr91, Thr93, and Ala95).

RESULTS AND DISCUSSION

Design of a Charge-Free Protein. To identify a low-charge protein with a known well-defined tertiary structure, we searched the Protein Data Bank (PDB) for a soluble protein of >100 residues that has the smallest number of titratable residues. Using this method, we identified a cellulose-binding domain from *Cellulomonas fimi*, xylanase B (PDB entry 1EXG⁴¹), with only four residues of 110 being charged, K28, D36, R68, and H90 (Figure 1). Three of these four charged residues are fully solvent exposed, while H90 is only slightly accessible (14% calculated with GETAREA⁴²). None are located at the cellulose-binding site.⁴³ To assess whether some of the charged residues might be essential for structure or function, we aligned the sequence with homologous protein domains from the NCBI database. This indicated that none of the four charges were conserved (Figure S1).

The all-β 3D structure of this protein domain, which we term EXG:CBM, appears to be quite conventional with a well-defined hydrophobic core, and a disulfide bond, but without any tyrosine or free cysteine residues (which are also titratable). Interestingly, however, the protein surface is dominated by hydroxyl amino acids rather than charged side chains.⁴¹ In addition, it is of practical relevance that tight binding to cellulose provides a simple assay for function as well as a method of purification.⁴⁴

Several noncharged substitutions of the four charged residues were chosen on the basis of the predicted changes in the free energy of unfolding (calculated by FoldX³⁶) and assessed experimentally (see Experimental Procedures). The mutant variant presented here, EXG:CBM^{MQMW}, carries the K28M, D36Q, R68M, and H90W mutations. Further chemical modification of the N-terminus using an acetylation reagent yielded a protein (acEXG:CBM^{MQMW}) that is entirely free of ionic charges at low pH (Figure 1).

Expression of EXG:CBM^{MQMW} in *E. coli* yielded approximately 8 mg of protein/L of culture. It should also be mentioned that EXG:CBM is part of a secreted protein with a single disulfide bond, and production in *E. coli* takes place via secretion into the periplasm. Whether the chaperone ensemble in the periplasm is more suited to deal with proteins with fewer charges, than the chaperones in the cytosol, is unknown.

EXG:CBM Remains Functional Despite Charge Depletion and Extreme pH Values. The binding of EXG:CBM to microcrystalline cellulose (Avicel) is strong ($K_d \approx 10^{-6}$ M^{43,45}), and on this basis, we have developed a purification protocol,³⁷ which includes a combination of batch coupling and elution from cellulose with hot distilled water (75–80 °C) followed by chromatography on Superdex 75. Following each step, pooled fractions were concentrated by lyophilization. This procedure was also implemented for the mutant protein, although the temperature at which the mutant protein was eluted from cellulose was somewhat lower (68–75 °C).

To assess the functionality of the EXG:CBM variants in a broad pH range, the proteins were dissolved in phosphate buffer at pH 2, 7, and 12, centrifuged to remove potential precipitation, and transferred to a new tube to which cellulose was added and collected after incubation for 1 h. EXG:CBM, EXG:CBM^{MQMW}, and acEXG:CBM^{MQMW} retained their binding to the cellulose under all pH conditions, resulting in complete pull down (Figure 2). These results are in agreement

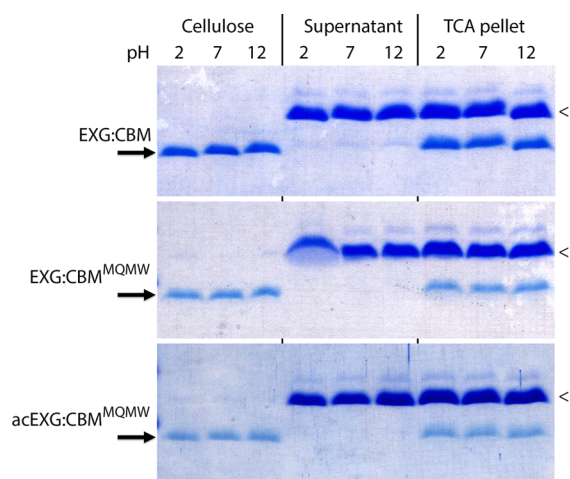


Figure 2. Cellulose binding assay. The indicated EXG:CBM protein variants and ribonuclease A (included as a loading control) were incubated with Avicel for 1 h followed by centrifugation to precipitate functional EXG:CBM bound to cellulose. “Cellulose” and “Supernatant” fractions were subjected to Tricine–SDS–PAGE. “TCA pellet” refers to TCA pellets from control samples without Avicel but otherwise treated in the same manner. The arrows indicate the position of EXG:CBM bands (2 μg per lane for EXG:CBM and EXG:CBM^{MQMW} and 4 μg per lane for acEXG:CBM^{MQMW}), while the higher-molecular weight bands (indicated with <) are ribonuclease A (2 μg per lane), which stays in solution over the full pH range.

with those of previous experiments showing the binding to be independent of pH in the range from 3 to 9.⁴⁴ The indifference toward pH for all three variants of EXG demonstrates that none of the charged residues are vital for the binding of EXG:CBM to cellulose. These observations are consistent with previous results identifying the cellulose-binding site to mainly comprise residues Trp17, Trp54, and Trp72 and a few other potential hydrogen bonding residues in the vicinity of these.⁴³

Charge Depletion of EXG:CBM Leaves the Protein Highly Soluble. Because previous studies report that charge depletion can inflict lowered protein solubility³⁰ and increased precipitation rates,³¹ we examined these properties of the EXG:CBM variants. Following incubation of the proteins at concentrations of 7–10 mg/mL for a period of up to 2 weeks, we did not observe any precipitation, and the proteins remained stable at pH 2, 7, and 12 (Figure 3 and Figure S2).

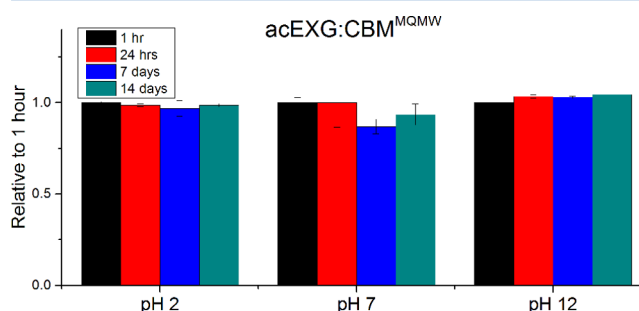


Figure 3. Precipitation propensity of acEXG:CBM^{MQMW}. The tendency of acEXG:CBM^{MQMW} to precipitate out of solution under three different pH conditions was followed over time by high-speed centrifugation and measurement of the absorbance at 280 nm. Measurements were performed in triplicate and normalized to the mean of the measurements at 1 h. The initial protein concentrations were 7–10 mg/mL.

Notably, even in the fully uncharged state, at pH 2, acEXG:CBM^{MQMW} remains soluble and retains its cellulose binding functionality completely after 2 weeks (Figure S3). As such, neither the substitutions themselves nor the charge removal seems to have a substantial impact on protein functionality.

One of the observations from the cellulose binding assay was that the wild type and the EXG:CBM charge variants are readily precipitated by TCA (Figure 2). On the other hand, no sign of precipitation was observed even after extended periods of time in 50 mM buffer (Figure 3). Still, we would expect the charge depletion to influence the equilibrium between the soluble and solid phase. All three proteins are highly soluble in MOPS buffer (≥ 55 mg/mL), so to assay differences in solubility, we used ammonium sulfate as a precipitant (Figure 4). The initial protein concentration did not have any impact on the measured solubility in the range of concentrations used for the experiments (Table S1, based on duplicate measurements).

As anticipated,¹⁴ solubility shows a close-to-logarithmic dependency on the ammonium sulfate concentration. While the slope is essentially the same, the solubility drops as a consequence of removing the four charged residues, with an additional decrease also related to removing the N-terminal charge, even though the latter actually increases the overall net charge of the protein (from 0 in EXG:CBM^{MQMW} to –1 in acEXG:CBM^{MQMW} at pH 7). Linear extrapolation of the data for EXG:CBM in Figure 4 to a salt-free solution results in an apparent solubility of ≈ 6 g/mL, but salting-in effects will probably level off the solubility at a lower ionic strength. Indeed, the data for EXG:CBM^{MQMW} and acEXG:CBM^{MQMW} do not seem to be strictly linear on a logarithmic scale at low salt concentrations. Experiments with polyethylene glycol (PEG) were conducted to obtain an extrapolated solubility in pure water, but we were not able to precipitate EXG:CBM with

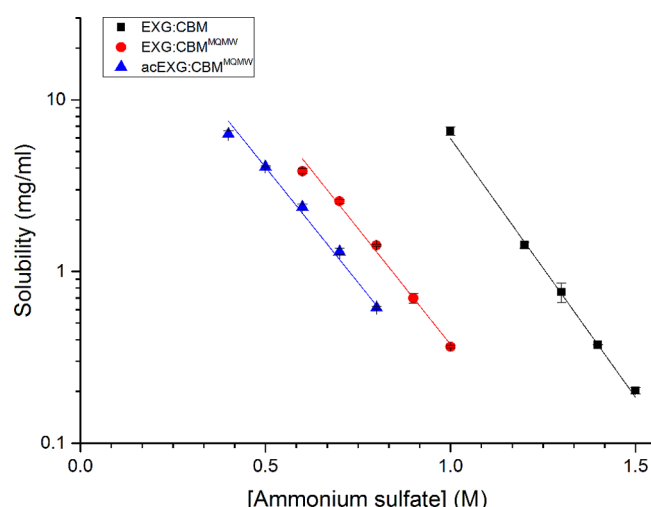


Figure 4. Ammonium sulfate precipitation of EXG:CBM, EXG:CBM^{MQMW}, and acEXG:CBM^{MQMW}. The protein solubility was measured in different concentrations of ammonium sulfate. The standard deviation of three independent measurements is shown. The lines show the best exponential fit. All measurements and preparations were performed at pH 7 and 25 °C.

up to 40% PEG 10000 and protein at concentrations as high as 20 mg/mL in 50 mM MOPS.

How does EXG:CBM maintain solubility in the absence of charges? First, an intramolecular disulfide bond tethers the N- and C-termini, making aggregation by unfolding from the termini difficult. Second, the surface of EXG:CBM is rich in hydroxy amino acids (Ser and Thr), which occupy a third of the solvent accessible surface area. This is ~ 3 times more than in other well-studied, globular proteins (Table S2). Also amide residues (Asn and Gln) are abundant on the surface (Table S2), and this could explain why even the charge-modified variants can maintain such an unexpected high solubility (≥ 55 mg/mL). Interestingly, the three soluble proteins (≥ 100 amino acids) in the PDB with only five or six charged residues (PDB entries 2K3O, 2KHM, and 3NDY) all have a similar abnormal abundance of hydroxy and amide amino acids on the surface (Table S2 and ref 46). This could also provide a plausible explanation for the low solubility observed for charge-free ubiquitin and protein S6.^{30,31} The surface of ubiquitin would be enriched with amide groups arising from the carbamylation of the positive residues but would still have a very low fraction of hydroxy groups compared to that of EXG:CBM. Even though all Lys and Arg residues are replaced with Ser in protein S6, the total surface area of hydroxy amino acids would still be only 16% in our simple analysis (Table S2). This is still only approximately half of what is seen in the natural weakly charged proteins. The effects of the introduction of hydroxy and amide residues on protein solubility have been tested experimentally,^{14,15} and although only a modest tendency to increase the solubility is seen, the introduction of such residues on a massive scale might effectively keep a protein soluble. If this is true, charge depletion of other weakly charged proteins, which contain a large fraction of hydroxy amino acids on their surface (Table S2), might also result in long-lived, soluble constructs.

The solubility of EXG:CBM^{MQMW} was significantly affected by removing the positive N-terminus (Figure 4). This appears to be in conflict with both the observation of the negligible effect of positive surface area on solubility¹⁶ and other studies showing that the positively charged amino and guanidino

groups of lysine and arginine residues, respectively, are not able to increase the solubility of hydrophobic patches in proteins.^{47,48} On the other hand, acEXG:CBM^{MQMW} has a very low net charge (-1 at pH 7), and lysine residues have been shown to increase the solubility of proteins with a low or positive net charge.^{14,15} Further, it is possible that the small number of charges unmasks effects that are difficult to study in proteins with a greater number of charged residues. The observation that the solubility of all EXG:CBM variants was affected by the presence of ammonium sulfate (Figure 4) supports the current view that ammonium sulfate precipitation does not work by screening electrostatic protein interactions but rather by excluding water from the protein surface.⁴⁹ Trichloroacetic acid (TCA) was also observed to efficiently precipitate both EXG:CBM and charge-modified variants (Figure 2) even though they are not sensitive to low pH. The mechanism of TCA precipitation is still not fully understood,⁵⁰ but our results indicate that the precipitating qualities of TCA are not solely linked to disruption of electrostatic interactions. Another result of practical relevance is the observation that the Coomassie staining efficiency of acEXG:CBM^{MQMW} is reduced several-fold relative to that of EXG:CBM (Figure 2). This is most likely caused by the removal of positive charges.^{51,52}

Stability and Structure Are Preserved in EXG:CBM^{MQMW} and acEXG:CBM^{MQMW}. The thermodynamic stability of EXG:CBM, EXG:CBM^{MQMW}, and acEXG:CBM^{MQMW} showed a cooperative denaturant-induced unfolding/refolding transition at both neutral pH and low pH (Figure 5 and Figure S5), which allowed us to determine

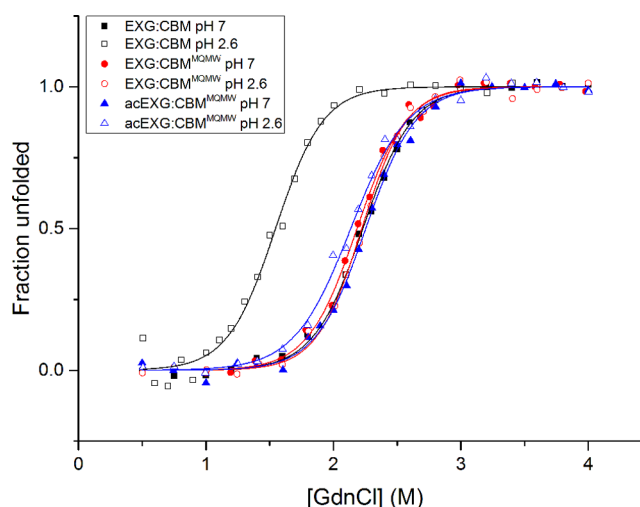


Figure 5. Chemical unfolding at neutral and low pH shows that stabilities are unaffected by charge state except for protonation of His90 in wild-type EXG:CBM. Protein unfolding was measured via the change in Trp fluorescence intensity. Solid lines represent fits using the linear extrapolation method⁵⁶ normalized to the fraction of unfolded protein. A non-normalized version of the figure can be seen in Figure S5.

ΔG_{D-N} , C_m , and m values (Table 1). The unfolding was in all cases fully reversible (Figure S4). We find that EXG:CBM is a relatively stable protein with an unfolding free energy at neutral pH (28 kJ mol^{-1}) that is comparable to those of many similarly sized proteins.⁵³

The stabilities of EXG:CBM^{MQMW} and acEXG:CBM^{MQMW} are within error the same as that of the parent protein at neutral

Table 1. Parameters for Guanidine Hydrochloride-Induced Unfolding

	EXG:CBM		EXG:CBM ^{MQMW}		acEXG:CBM ^{MQMW}	
	pH 2.6	pH 7	pH 2.6	pH 7	pH 2.6	pH 7
m (kJ mol ⁻¹ M ⁻¹) ^a	13.4 ± 1.6	12.7 ± 0.7	14.1 ± 2.0	12.9 ± 0.5	11.8 ± 2.7	13.0 ± 1.1
C_m (M) ^b	1.54 ± 0.03	2.23 ± 0.04	2.22 ± 0.03	2.18 ± 0.00	2.14 ± 0.03	2.25 ± 0.00
ΔG_{D-N} (kJ mol ⁻¹) ^a	20.7 ± 2.6	28.3 ± 1.1	31.3 ± 4.8	28.2 ± 1.0	25.2 ± 5.8	29.3 ± 2.5

^aThe m value and ΔG_{D-N} were both obtained by fitting to the linear extrapolation method (see [Experimental Procedures](#)). The values presented here are the averages of several experiments, and their uncertainties represent the maximal deviations of the results of individual experiments from this average rather than the uncertainty from global fitting of the entire data set. ^b C_m was calculated from m and ΔG_{D-N} .

pH, demonstrating that charge removal does not change the overall protein stability. Lowering the pH from 7.0 to 2.6 causes a decrease in the stability of EXG:CBM by 8 ± 3 kJ mol⁻¹, suggesting that the protonation of one or more residues [most likely H90 (see below)] decreases stability. In agreement with this suggestion, no such decrease was observed for EXG:CBM^{MQMW} or acEXG:CBM^{MQMW}. The average magnitude of the m value (13 kJ mol⁻¹ M⁻¹) matches the expected value (13.4 ± 0.7 kJ mol⁻¹ M⁻¹)⁵⁴ from unfolding a protein of 105 residues, the number of ordered residues in EXG:CBM. Thus, despite the unusual sequence composition and even after charge removal, the protein behaves like other proteins of the same size.

To study the impact of pH on protein structure more directly, we recorded a series of two-dimensional ¹H–¹⁵N HSQC NMR spectra as a function of pH for both EXG:CBM and EXG:CBM^{MQMW}. The spectra show that both proteins retain very similar three-dimensional structures over a broad pH interval ([Figure 6](#) and [Figures S6 and S7](#)). Far- and near-

studying protonation states in the EXG:CBM system. As NMR has previously been used to determine the structure of EXG:CBM,⁴¹ assignments of the wild-type HSQC spectra could be adopted directly. The five N-terminal amino acids of the protein cannot be detected in the spectra and are thought to be unstructured.⁴¹ After assignment of the peaks, titration of EXG:CBM was conducted from pH 2 to 9 in increments of 1/3 pH unit. The protein clearly remains structured over the entire pH interval with only minor changes in the chemical shifts for most residues ([Figure S6](#)). However, the signal for H90 and several peaks from residues spatially close disappear between pH 3 and 5, suggesting a conformational change or local adaptation to the protonation of H90 taking place with an intermediate exchange rate. Despite the small number of titratable residues in EXG:CBM, we were able to detect significant changes in more than half of the backbone amide chemical shifts. The simplicity of the system is, however, evident from the fact that all titration curves could be fitted well by a single Henderson–Hasselbalch equation ([Figure S9](#)). The pK_a values of D36, H90, and the C-terminal G110 were found to be 4.2, 4.0, and 3.5, respectively. The measured pK_a value of H90 (4.0) is substantially shifted downward relative to the reference value (6.5) in an alanine pentapeptide,⁵⁵ suggesting that the protonation of this residue could be responsible for the destabilization of EXG:CBM observed at low pH. This is further supported by the fact that the drop in stability is not observed in EXG:CBM^{MQMW} or other mutants not containing the His residue. Because no chemical shift changes were observed above pH 6 for EXG:CBM ([Figure S6](#)), the titration of EXG:CBM^{MQMW} was conducted from pH 1.8 to 6.2. The resulting spectra clearly demonstrated the persistence of the protein toward changes in pH ([Figure 6](#)). Only very few changes in chemical shifts are evident, all of which fitted to the Henderson–Hasselbalch equation and is consistent with titration of the C-terminus with a pK_a of 3.7 ± 0.2 ([Figure S10](#)). The inertness of the protein to pH might in part be sustained by its unstructured N-terminus, giving the N-terminal charge only little uniform effect on the rest of the protein (much like a free ion). In fact, not a single peak in the EXG:CBM HSQC experiments ([Figure S9](#)) is significantly affected by the presumed protonation of the N-terminus ($pK_a \sim 8$) in the pH interval between 6 and 9. Taken together, these data clearly indicate that changing the pH has an only limited impact on the structure of EXG:CBM^{MQMW}.

CONCLUDING REMARKS

We have demonstrated that the charge-minimized protein EXG:CBM^{MQMW} is folded and functional as judged by several criteria. (i) It is capable of high-affinity cellulose binding. (ii) It has a well-defined folding transition similar in shape and stability to that of the wild type. (iii) NMR spectra are well dispersed and similar to that of the wild type. (iv) Its solubility

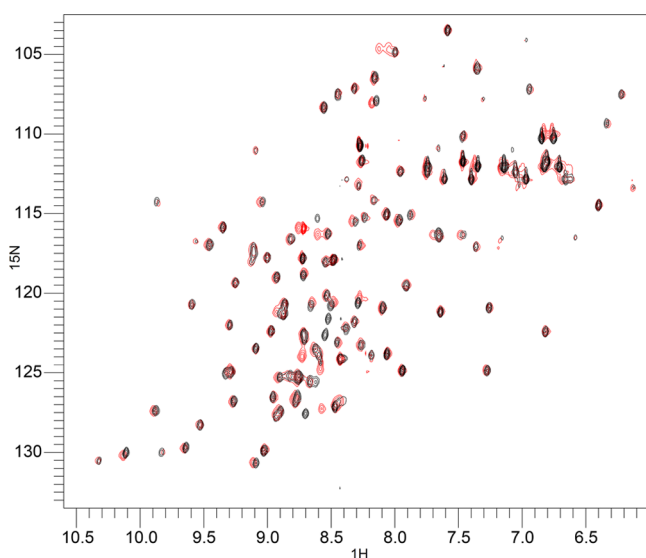


Figure 6. EXG:CBM^{MQMW} ¹⁵N HMQC spectra at pH 1.8 (red) and pH 6.3 (black).

UV circular dichroism (CD) data of acEXG:CBM^{MQMW} at pH 2, 7, and 11 indicate that the secondary and tertiary structure of this variant is also unperturbed by the change in pH ([Figure S8](#)). The lack of pH sensitivity is unusual but is consistent with the notion that the charged residues of EXG are nonessential for the stability and function of this protein.

Effect of pH on Wild-Type and Charge-Neutral EXG:CBM Studied by NMR. Because the complexity of charge interactions is highly reduced, we were interested in

is surprisingly high (at least 55 mg/mL), and it does not have any tendency to precipitate over extended periods of time. These results also apply to the acetylated form of the mutant protein. While previous reports of charge-free proteins depend heavily on low pH to be uncharged,^{30,31} our study is, to the best of our knowledge, the first to show a fully functional protein completely free of titratable side chains. It demonstrates that while charged residues in many cases do aid the *in vivo* folding and solubility of proteins, they are not an absolute prerequisite. The existence of charge-depleted EXG:CBM^{MQMW} therefore adds to our understanding of what is achievable with protein design, and we envisage that this protein can be useful in helping elucidate detailed electrostatic properties in proteins.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.6b00269.

Multiple-sequence alignment of family 2 CBMs, precipitation propensity of EXG:CBM and EXG:CBM^{MQMW}, cellulose binding assay, guanidine hydrochloride unfolding curves, far- and near-UV CD of acEXG:CBM^{MQMW}, ¹H–¹⁵N HSQC spectra of EXG:CBM, titration curves, and Tables S1 and S2 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +45 3532 1500. E-mail: jrwinther@bio.ku.dk.

Present Addresses

[†]C.K.: Center for Evolutionary Chemical Biology, Department of Chemistry, University of Copenhagen, DK-2200 Copenhagen N, Denmark.

[‡]R.E.: Section for Protein Sciences, Department of Systems Biology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

[§]K.E.J.: Section of Pharmaceutical Technology and Engineering, Department of Pharmacy, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark.

^{||}M.V.: Institute of Theoretical Physics, University of Cologne, 50937 Köln, Germany.

Author Contributions

C.H., C.K., and R.E. contributed equally to this work.

Funding

The project was funded in part by Grant NNF12OC0000793 from the Novonordisk foundation to J.R.W. and Grant 09-070508 from the Danish Council for Independent Research, Technology and Production Sciences (FTP) to M.W.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

CD, circular dichroism; PDB, Protein Data Bank; PEG, polyethylene glycol; TCA, trichloroacetic acid.

■ ADDITIONAL NOTES

^aWe will throughout the text use the term “charged residues” to refer to the titratable residues.

^bUniProtKB/Swiss-Prot protein knowledgebase release 2015_09 statistics.

■ REFERENCES

- (1) Sheinerman, F. B., Norel, R., and Honig, B. (2000) Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10, 153–159.
- (2) Brock, K., Talley, K., Coley, K., Kundrotas, P., and Alexov, E. (2007) Optimization of electrostatic interactions in protein-protein complexes. *Biophys. J.* 93, 3340–3352.
- (3) Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007) Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* 129, 10110–10112.
- (4) Miklos, A. E., Kluwe, C., Der, B. S., Pai, S., Sircar, A., Hughes, R. A., Berrondo, M., Xu, J., Codrea, V., Buckley, P. E., Calm, A. M., Welsh, H. S., Warner, C. R., Zacharko, M. A., Carney, J. P., Gray, J. J., Georgiou, G., Kuhlman, B., and Ellington, A. D. (2012) Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* 19, 449–455.
- (5) Simeonov, P., Berger-Hoffmann, R., Hoffmann, R., Strater, N., and Zuchner, T. (2011) Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Eng. Des. Sel.* 24, 261–268.
- (6) Xiao, L., and Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* 289, 1435–1444.
- (7) Zhou, X. X., Wang, Y. B., Pan, Y. J., and Li, W. F. (2008) Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids* 34, 25–33.
- (8) Yokota, K., Satou, K., and Ohki, S. (2006) Comparative analysis of protein thermo stability: Differences in amino acid content and substitution at the surfaces and in the core regions of thermophilic and mesophilic proteins. *Sci. Technol. Adv. Mater.* 7, 255–262.
- (9) Jeppesen, M. D., Westh, P., and Otzen, D. E. (2010) The role of protonation in protein fibrillation. *FEBS Lett.* 584, 780–784.
- (10) Morshedi, D., Ebrahim-Habibi, A., Moosavi-Movahedi, A. A., and Nemat-Gorgani, M. (2010) Chemical modification of lysine residues in lysozyme may dramatically influence its amyloid fibrillation. *Biochim. Biophys. Acta, Proteins Proteomics* 1804, 714–722.
- (11) Es-haghi, A., Shariatizi, S., Ebrahim-Habibi, A., and Nemat-Gorgani, M. (2012) Amyloid fibrillation in native and chemically-modified forms of carbonic anhydrase II: Role of surface hydrophobicity. *Biochim. Biophys. Acta, Proteins Proteomics* 1824, 468–477.
- (12) Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805–808.
- (13) Shaw, B. F., Moustakas, D. T., Whitelegge, J. P., and Faull, K. F. (2010) Taking Charge of Proteins: From Neurodegeneration to Industrial Biotechnology. *Adv. Protein Chem. Struct. Biol.* 79, 127–164.
- (14) Trevino, S. R., Scholtz, J. M., and Pace, C. N. (2007) Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J. Mol. Biol.* 366, 449–460.
- (15) Islam, M. M., Khan, M. A., and Kuroda, Y. (2012) Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant. *Biochim. Biophys. Acta, Proteins Proteomics* 1824, 1144–1150.
- (16) Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N., and Scholtz, J. M. (2012) Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophys. J.* 102, 1907–1915.
- (17) Collins, K. D. (1997) Charge density-dependent strength of hydration and biological structure. *Biophys. J.* 72, 65–76.
- (18) Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* 216, 1031–1044.
- (19) Xiao, S., Patsalo, V., Shan, B., Bi, Y., Green, D. F., and Raleigh, D. P. (2013) Rational modification of protein stability by targeting surface sites leads to complicated results. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11337–11342.
- (20) Sali, D., Bycroft, M., and Fersht, A. R. (1991) Surface electrostatic interactions contribute little of stability of barnase. *J. Mol. Biol.* 220, 779–788.

- (21) Gribenko, A. V., Patel, M. M., Liu, J., McCallum, S. A., Wang, C., and Makhatadze, G. I. (2009) Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2601–2606.
- (22) Arnold, U., and Schopfel, M. (2012) Significant stabilization of ribonuclease A by additive effects. *FEBS J.* 279, 2508–2519.
- (23) Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (2000) Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* 39, 12375–12381.
- (24) Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.* 327, 1135–1148.
- (25) Donald, J. E., Kulp, D. W., and DeGrado, W. F. (2011) Salt bridges: geometrically specific, designable interactions. *Proteins: Struct., Funct., Genet.* 79, 898–915.
- (26) Webb, H., Tynan-Connolly, B. M., Lee, G. M., Farrell, D., O'Meara, F., Sondergaard, C. R., Teilum, K., Hewage, C., McIntosh, L. P., and Nielsen, J. E. (2011) Remeasuring HEWL pK(a) values by NMR spectroscopy: methods, analysis, accuracy, and implications for theoretical pK(a) calculations. *Proteins: Struct., Funct., Genet.* 79, 685–702.
- (27) Farrell, D., Miranda, E. S., Webb, H., Georgi, N., Crowley, P. B., McIntosh, L. P., and Nielsen, J. E. (2010) Titration_DB: storage and analysis of NMR-monitored protein pH titration curves. *Proteins: Struct., Funct., Genet.* 78, 843–857.
- (28) Gosink, L. J., Hogan, E. A., Pulsipher, T. C., and Baker, N. A. (2014) Bayesian model aggregation for ensemble-based estimates of protein pK(a) values. *Proteins: Struct., Funct., Genet.* 82, 354–363.
- (29) Nielsen, J. E., Gunner, M. R., and García-Moreno, E. B. (2011) The pK(a) Cooperative: A collaborative effort to advance structure-based calculations of pK(a) values and electrostatic effects in proteins. *Proteins: Struct., Funct., Genet.* 79, 3249–3259.
- (30) Loladze, V. V., and Makhatadze, G. I. (2002) Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable. *Protein Sci.* 11, 174–177.
- (31) Kurnik, M., Hedberg, L., Danielsson, J., and Oliveberg, M. (2012) Folding without charges. *Proc. Natl. Acad. Sci. U. S. A.* 109, 5705–5710.
- (32) Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262, 1680–1685.
- (33) Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., and Baker, D. (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* 4, 805–809.
- (34) Akanuma, S., Kigawa, T., and Yokoyama, S. (2002) Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13549–13553.
- (35) Walter, K. U., Vamvaca, K., and Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* 280, 37742–37746.
- (36) Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–388.
- (37) Knudsen, K. B., Kofoed, C., Espersen, R., Hojgaard, C., Winther, J. R., Willemoes, M., Wedin, I., Nuopponen, M., Vilske, S., Aimonen, K., Weydahl, I. E., Alenius, H., Norppa, H., Wolff, H., Wallin, H., and Vogel, U. (2015) Visualization of Nanofibrillar Cellulose in Biological Tissues Using a Biotinylated Carbohydrate Binding Module of beta-1,4-Glycanase. *Chem. Res. Toxicol.* 28, 1627–1635.
- (38) Schagger, H. (2006) Tricine-SDS-PAGE. *Nat. Protoc.* 1, 16–22.
- (39) Walker, J. M. (2005) *The proteomics protocols handbook*, Humana Press, Totowa, NJ.
- (40) Lian, L.-Y., and Roberts, G. C. K. (2011) *Protein NMR spectroscopy: practical techniques and applications*, Wiley, Chichester, U.K.
- (41) Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harrisbrandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) Solution Structure of a Cellulose-Binding Domain from *Cellulomonas-Fimi* by Nuclear-Magnetic-Resonance Spectroscopy. *Biochemistry* 34, 6993–7009.
- (42) Fraczekiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* 19, 319–333.
- (43) McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000) Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng., Des. Sel.* 13, 801–809.
- (44) Ong, E., Gilkes, N. R., Miller, R. C., Warren, R. A. J., and Kilburn, D. G. (1993) The Cellulose-Binding Domain (Cbdcx) of an Exoglucanase from *Cellulomonas-Fimi* - Production in *Escherichia-Coli* and Characterization of the Polypeptide. *Biotechnol. Bioeng.* 42, 401–409.
- (45) Jervis, E. J., Haynes, C. A., and Kilburn, D. G. (1997) Surface diffusion of cellulases and their isolated binding domains on cellulose. *J. Biol. Chem.* 272, 24016–24023.
- (46) Hagn, F., Eisoldt, L., Hardy, J. G., Vendrely, C., Coles, M., Scheibel, T., and Kessler, H. (2010) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* 465, 239–U131.
- (47) Chan, P., Curtis, R. A., and Warwicker, J. (2013) Soluble expression of proteins correlates with a lack of positively-charged surface. *Sci. Rep.* 3, 3333 DOI: 10.1038/srep03333.
- (48) Yagi, S., Akanuma, S., and Yamagishi, A. (2014) Addition of negatively charged residues can reverse the decrease in the solubility of an acidic protein caused by an artificially introduced non-polar surface patch. *Biochim. Biophys. Acta, Proteins Proteomics* 1844, 553–560.
- (49) Wingfield, P. (2001) Protein Precipitation Using Ammonium Sulfate. In *Current Protocols in Protein Science*, John Wiley & Sons, Inc., New York.
- (50) Rajalingam, D., Loftis, C., Xu, J. J., and Kumar, T. K. (2009) Trichloroacetic acid-induced protein precipitation involves the reversible association of a stable partially structured intermediate. *Protein Sci.* 18, 980–993.
- (51) de Moreno, M. R., Smith, J. F., and Smith, R. V. (1986) Mechanism studies of coomassie blue and silver staining of proteins. *J. Pharm. Sci.* 75, 907–911.
- (52) Tal, M., Silberstein, A., and Nusser, E. (1985) Why Does Coomassie Brilliant Blue-R Interact Differently with Different Proteins - a Partial Answer. *J. Biol. Chem.* 260, 9976–9980.
- (53) Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J. C., Bona, D., Miller, E. J., Vallee-Belisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N. D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., Kragelund, B. B., Michnick, S. W., Chiti, F., Bai, Y., Hagen, S. J., Serrano, L., Oliveberg, M., Raleigh, D. P., Wittung-Stafshede, P., Radford, S. E., Jackson, S. E., Sosnick, T. R., Marqusee, S., Davidson, A. R., and Plaxco, K. W. (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* 14, 602–616.
- (54) Geierhaas, C. D., Nickson, A. A., Lindorff-Larsen, K., Clarke, J., and Vendruscolo, M. (2006) BPPred: a Web-based computational tool for predicting biophysical parameters of proteins. *Protein Sci.* 16, 125–134.
- (55) Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2006) pK values of the ionizable groups of proteins. *Protein Sci.* 15, 1214–1218.
- (56) Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27, 8063–8068.